DOCUMENT RESUME

ED 275 731 TM 860 645

AUTHOR McIntyre, Kenneth E.
TITLE Using Classroom Observation Data for Diagnostic Purposes.
PUB DATE Apr 86
NOTE 12p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Algebra; *Classroom Observation Techniques; Classroom Research; *Evaluation Methods; *Formative Evaluation; Interrater Reliability; Language Arts; Models; Scoring; Secondary Education; *Teacher Evaluation
IDENTIFIERS Data Envelopment Analysis

ABSTRACT

This paper dealt with the use of classroom observation data for formative evaluation purposes, and with a research project in which scores based on observed performance of teachers in secondary school algebra and English classes were compared with efficiency scores based on an input-output model. The model, using Data Envelopment Analysis (DEA) and Constrained Facet Analysis (CFA), was developed by the Educational Productivity Council at the University of Texas at Austin. DEA/CFA was used previously to compare the outputs of schools or school districts that have similar inputs. This paper reported on the use of the model at the individual teacher level. An observation guide, developed from the process-product research, contained 39 indicators in 7 categories. Three trained observers visited 140 classrooms for a full period each visit. A comparison of the teachers' scores based on observed performance with their efficiency scores using the DEA/CFA input-output model revealed a high correlation for the algebra classes but no significant correlation for the English classes. (Author/JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

FD275731

Using Classroom Observation Data for Diagnostic Purposes

Kenneth E. McIntyre, University of Texas at Austin

AERA Symposium, San Francisco, 1986

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☒ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. McIntyre

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 860 645

ABSTRACT

"Using Classroom Observation Data for Diagnostic Purposes",
Kenneth E. McIntyre, University of Texas at Austin,
AERA Symposium, San Francisco, 1986

This paper dealt with the use of classroom observation data for formative evaluation purposes, and with a research project in which scores based on observed performance of teachers in secondary school Algebra and English classes were compared with efficiency scores based on an input-output model. The model, using Data Envelopment Analysis (DEA) and Constrained Facet Analysis (CFA), was developed by the Educational Productivity Council at the University of Texas at Austin. DEA/CFA was used previously to compare the outputs of schools or school districts that have similar inputs. This paper reported on the use of the model at the individual teacher level.

An observation guide, developed from the process-product research, contained 39 indicators in seven categories. Three trained observers visited 140 classrooms for a full period each visit. A comparison of the teachers' scores based on observed performance with their efficiency scores using the DEA/CFA input-output model revealed a high correlation for the Algebra classes but no significant correlation for the English classes.

The use of both approaches -- classroom observation and DEA/CFA -- for diagnostic purposes to improve instruction was discussed.

Kenneth E. McIntyre
University of Texas, Austin

AERA Symposium
"Using Classroom Observation Data for Diagnostic Purposes"

The observation of teacher performance in the classroom is one way of collecting data concerning teachers for both formative and summative purposes, as well as for research on the efficiency or effectiveness of schools or individual teachers. This presentation will deal with the use of observation of instruction in the classroom to collect data to relate to teacher efficiency as measured by an input-output model that will be discussed later in the symposium.

Classroom observation, if it is competently done, can make a significant and valuable contribution to the data-collection process for many worth-while purposes. However, as with any other complex activity, it has its limitations, which must be recognized and dealt with in a realistic way:

(1) The observation instrument that is used must be sound, focusing on teacher and student behavior that is relevant to the accomplishment of appropriate and important objectives of the organization. The instrument should be descriptive and non-judgmental in nature, requiring the observer to record what happens rather than judgments concerning what happens.

(2) The observer must be thoroughly trained to use the instrument. Observation training should produce relatively high levels of inter-rater reliability.

(3) Observations must be scheduled to take into account the differences in teacher performance from class to class and from situation to situation during the year, the week, and even the

day. Generalizations concerning teacher behavior are hazardous, unless based upon several samples of that behavior in different situations. Observers are, at best, almost always limited to a few relatively brief observation periods in a given teacher's classroom. This creates the possibility that apparent imbalances in the teacher's emphases on different behaviors are to some extent the result of inadequate sampling of the teacher's overall performance.

(4) Keeping the preceding cautions in mind, the process-product research on teacher performance justifies the development of generalizations that can be used to support conclusions and decisions concerning teacher efficiency and effectiveness. For example, the observation instrument that was developed for, and used in, the project to be discussed here contains such generically useful items as the following:

Clearly reveals aims, objectives, and a plan for the lesson.

Engages students in active (vs. passive) learning.

Utilizes various means of rewarding desirable behavior or responses.

Shows excitement or interest in the lesson, through verbal or non-verbal means.

Maintains the classroom as a businesslike, orderly work place.

Engages in formative evaluation through monitoring, asking questions (oral/written), feedback, and reteaching if needed.

Although such evidences as the above are almost always desirable, even the most obviously "good" behaviors in education

or in life in general can be carried to such extremes that they are no longer good. Eating is necessary, but one can do too much of it; likewise, too much sleep, too much exercise, too much talk—all of which are good in moderation—can be carried to extremes that are dysfunctional. Even honesty, which we all know is "the best policy," is not always advisable; for example, who would advocate telling a mother that her baby is ugly—even if the baby obviously is ugly?

The point is that the users of observation instruments must recognize that more is not necessarily better, with regard to several of the indicators that could legitimately be included in such instruments. For example, the individualization of instruction, in view of the wide range in aptitudes and abilities of students, is generally a desirable activity of teachers. However, individualized instruction is not always the most effective means of meeting students' needs, and it is more appropriate for some students than it is for others. Similarly, it is generally desirable for teachers to include questions at higher cognitive levels than simple rote memory, but questions at the recognition or recall levels are sometimes useful. Rewarding students for good behavior or correct responses is generally to be encouraged; yet praise should be earned and sincerely given, and even praise can be so excessive that it becomes ineffective.

These caveats place a burden on classroom observers to be keenly aware of the subtleties inherent in such complex activities as teaching and learning. Merely recording or even counting evidences of a particular teaching behavior, and drawing conclusions that are based on a "more is better" philosophy, has

merit in general and up to a point is a useful procedure with regard to many observable behaviors. However, observers must be sensitive to the exceptions. Having access to relevant information will help--such as the teacher's plans and objectives for the lesson, copies of teacher-made tests, and data concerning students in the class. Sensing when a point of diminishing returns on one indicator has been reached in a class is sometimes abetted by noting related evidence on another indicator; for example, although teacher modeling is often a desirable and useful practice, too much modeling might be reflected in a decline in student interest and an increase in misbehavior.

The observation instrument that we developed at The University of Texas was designed for use in academic classrooms at the secondary school level, although many of the items would be relevant in classrooms from the elementary level through college. It is a comprehensive guide, including an array of indicators of effective performance. The principle of parsimony was used in developing the guide, in that only those indicators supported by reputable process-product research and/or expert opinion were included--a total of thirty-nine indicators in seven categories. Based on our experience with the guide during the project, it was revised to its present form, which includes thirty-three indicators in six categories. The indicators are all stated positively, although any kind of relevant evidence is recorded.

The format of the observation instrument was influenced by a desire to enhance accuracy, efficiency, reliability, and even convenience. A common complaint of observers is that their observation instruments are so long and unwieldy that they spend

too much time shuffling pages in order to locate an unmanageable number of items, so that they cannot concentrate on observing and recording important matters. Our present instrument consists of one page--an 8 1/2 x 14-inch sheet, on which are listed the thirty-three indicators in the six categories--which the observer places at the top of the table and reviews frequently during the observation. There is a second sheet, divided into six areas for writing notes concerning the six categories. No paper shuffling needs to take place at all--yet the indicators are always on display to remind the observer to be alert to relevant incidents that occur.

After each observation visit has ended, the observer leaves the classroom and goes to an office or other room where the remainder of the task is completed in privacy. First, the observer reviews all of the notations, to ensure clarity and to check once more to be sure that no relevant evidence has been omitted. Then the observer rates each of the thirty-three indicators, on a seven-point scale, in terms of the amount of evidence recorded. The number 3 is circled for an indicator if no relevant evidence was observed--positive or negative. The number 2 is circled if the recorded evidence tends toward the negative side, and the 1 is circled if the evidence is strongly on the negative side (usually two or more significant examples). From 4 through 7, the circled numbers indicate increasing amounts of positive evidence concerning the indicator. Although observers are instructed not to be slaves to quantity, to the neglect of the importance and impact of each behavior, a general guide-line is to use a 4 as a rating for one significant behavior (or two less

significant ones), a 5 for two significant behaviors, a 6 for three, and a 7 for four or more. In the project discussed here, the observers visited four classrooms together (for a full period each time) before the project started, comparing notes and ratings after each visit, to reduce the variability in ratings to an acceptable level--which was achieved. Obviously, an element of judgment enters into this process of attaching numbers to behavior, even if the observers are trained and the instrument is constructed to encourage objectivity. However, it is common knowledge that relatively high levels of inter-rater reliability can be achieved when the raters are skilled in the use of a sound instrument.

Each of the three observers in the project made two full-period, scheduled visits to the classrooms of twenty-four teachers, with the exception of a few instances in which the teacher was absent on the scheduled day. A total of approximately 140 visits were made, involving middle school and high school teachers of either language arts or algebra.

The final step in the process followed by the three observers in the project, or for anyone else who uses the same instrument, is for the observer to calculate the mean for each category and then to add the means to arrive at a total score for the observation. Each of the means and the total score can subsequently be related to other measures, such as the efficiency scores of the teachers, which will be discussed by Dr. Linda Reaves later. A brief description of how the efficiency scores can be used will be provided here.

Data Envelopment Analysis (DEA) and Constrained Facet Analysis (CFA) are quite sophisticated processes that employ an input-output model to determine the efficiency of schools, whereby each school's outputs are compared with those of other schools having similar inputs. The fairness of having a school's efficiency determined by comparing its outputs with those of other schools with like inputs is obvious. After having used the DEA/CFA approach very successfully at the school level, the Educational Productivity Council at The University of Texas at Austin decided to extend the model to the classroom level, by comparing outputs of individual teachers with the outputs of other teachers having similar inputs.

The use of DEA/CFA along with the analysis of classroom observation can produce more useful information than either process alone can provide. In using a comprehensive observation guide that contains many indicators of quality teaching in general, one must realize that even a teacher who is efficient in comparison with other teachers via DEA/CFA analysis might have one or more weaknesses that can be corrected. For example, a teacher might be doing an exemplary job of teaching that which is tested and is used as a product measure, and hence has a high efficiency score, but the teacher might not be doing so well on other important behaviors that are observable in the classroom. This places a burden on both the observation instrument/process and the efficiency measures/process, to include as many <u>really</u> important and measurable variables as possible and practicable, without unduly loading the system with trivia. Likewise, it is conceivable that a teacher might be doing many of the "right

things" in the classroom, and yet have a lower-than-expected efficiency score, perhaps due to some of the situational variables mentioned earlier. For one thing, we should be sure that the efficiency score is based on the same class and the same circumstances with which the observation deals.

Both observation and efficiency measures can be used diagnostically to help individual teachers to improve. The observation instrument itself should, in the hands of a skilled observer, provide much useful data. Although, as stated earlier, not every indicator in the guide would necessarily be appropriate for every classroom situation, all of the indicators should be considered to be desirable unless the situation specifically alters the diagnosis. The observation guide, then, should serve as a "teacher helper"—not to encourage teachers to "put on a show" for observers whenever one is present, but to suggest ways in which even the best teachers might enhance their instruction and get to be even better. I am not worried about teachers' "putting on a show" if they know when an observer will be present or what instrument the observer will be using. The observer's main concern should be to learn what the teacher can do under favorable circumstances, and it is the instructional leader's job to make those circumstances favorable. It is true that teachers can and probably will do certain things if they know that the observer will want to see those things—but why not, if "those things" are important and are really performed well? Then, too, not many teachers can fake the essence of really good teaching practices; for example, if a teacher can demonstrate the use of analysis and synthesis questions through questioning strategies

and techniques that are discussed in the best professional literature, then the observer should be sophisticated enough to know that the teacher is not just going through the motions.

The efficiency analysis (DEA/CFA) that can be used as part of the process of determining and improving efficiency of a teacher should also be diagnostic in nature. Any time a teacher is found to be less than efficient, in comparison with other teachers having similar inputs, the DEA/CFA process can provide useful data for helping that teacher. For example, the teacher might be spending too much time on unproductive processes or resources, at the expense of more productive ones, and DEA/CFA can reveal this.

When the classroom observation data and the efficiency data are consistent, as they were found to be for the algebra teachers in the research project discussed here, this is reassuring to the instructional leader who is exploring ways to help the teacher involved. When the observation and efficiency data are not consistent, then explanations for the inconsistency can still be sought--and often found, as discussed earlier. In such cases, if it is the observation data that suggest the problem, then the observation instrument should be especially useful for diagnosis and planning purposes, and vice versa for situations of the opposite type. Both sets of data should be of value, however, in almost all conceivable situations.